

(19)

Europäisches Patentamt

European Patent Office

Office européen des brevets



(11)

EP 1 033 405 A2

(12)

## EUROPEAN PATENT APPLICATION

(43) Date of publication:  
06.09.2000 Bulletin 2000/36(51) Int Cl.7: C12N 15/29, C12N 15/82,  
C07K 14/415, C12Q 1/68,  
A01H 5/00

(21) Application number: 00301439.6

(22) Date of filing: 25.02.2000

(84) Designated Contracting States:  
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU  
MC NL PT SE  
Designated Extension States:  
AL LT LV MK RO SI

(30) Priority: 25.02.1999 US 121825 P  
27.07.1999 US 145918 P  
28.07.1999 US 145951 P  
02.08.1999 US 146388 P  
02.08.1999 US 146389 P  
02.08.1999 US 146386 P  
03.08.1999 US 147038 P  
04.08.1999 US 147302 P  
04.08.1999 US 147204 P

More priorities on the following pages

(83) Declaration under Rule 28(4) EPC (expert solution)

(71) Applicant: Ceres Incorporated  
Malibu, CA 90265 (US)

(72) Inventors:

- Alexandrov, Nickolai  
Thousand Oaks, CA 91320 (US)

- Brover, Vyacheslav  
Calabasas, CA 91302 (US)
- Chen, Xianfeng  
Los Angeles, CA 90025 (US)
- Subramanian, Gopalakrishnan  
Moorpark, CA 93021 (US)
- Troukhan, Maxim E.  
South Pasadena, CA 91030 (US)
- Zheng, Liansheng  
Creve Coeur, MO 63141 (US)
- Dumas, J.  
, (US)

(74) Representative:  
Bannerman, David Gardner et al  
Withers & Rogers,  
Goldings House,  
2 Hays Lane  
London SE1 2HW (GB)

## Remarks:

THE COMPLETE DOCUMENT INCLUDING  
REFERENCE TABLES AND THE SEQUENCE  
LISTING IS AVAILABLE ON CD-ROM FROM THE  
EUROPEAN PATENT OFFICE, VIENNA  
SUB-OFFICE.

(54) Sequence-determined DNA fragments and corresponding polypeptides encoded thereby

(57) The present invention provides DNA molecules that constitute fragments of the genome of a plant, and polypeptides encoded thereby. The DNA molecules are useful for specifying a gene product in cells, either as a promoter or as a protein coding sequence or as an UTR or as a 3' termination sequence, and are also useful in controlling the behavior of a gene in the chromosome,

in controlling the expression of a gene or as tools for genetic mapping, recognizing or isolating identical or related DNA fragments, or identification of a particular individual organism, or for clustering of a group of organisms with a common trait.

<sup>0</sup>Arabidopsis DNA is used in the present experiment, but the procedure is a general one.

EP 1 033 405 A2

**Description****FIELD OF THE INVENTION**

- 5 [0001] The present invention relates to isolated polynucleotides that represent a complete gene, or a fragment thereof, that is expressed. In addition, the present invention relates to the polypeptide or protein corresponding to the coding sequence of these polynucleotides. The present invention also relates to isolated polynucleotides that represent regulatory regions of genes. The present invention also relates to isolated polynucleotides that represent untranslated regions of genes. The present invention further relates to the use of these isolated polynucleotides and polypeptides and proteins.
- 10

**DESCRIPTION OF THE RELATED ART**

- 15 [0002] Efforts to map and sequence the genome of a number of organisms are in progress; a few complete genome sequences, for example those of *E. coli* and *Saccharomyces cerevisiae* are known (Blattner et al., *Science* 277:1453 (1997); Goffeau et al., *Science* 274:546 (1996)). The complete genome of a multicellular organism, *C. elegans*, has also been sequenced (See, the *C. elegans* Sequencing Consortium, *Science* 282:2012 (1998)). To date, no complete genome of a plant has been sequenced, nor has a complete cDNA complement of any plant been sequenced.

**20 SUMMARY OF THE INVENTION**

- 25 [0003] The present invention comprises polynucleotides, such as complete cDNA sequences and/or sequences of genomic DNA encompassing complete genes, fragments of genes, and/or regulatory elements of genes and/or regions with other functions and/or intergenic regions, hereinafter collectively referred to as Sequence-Determined DNA Fragments (SDFs), from different plant species, particularly corn, wheat, soybean, rice and *Arabidopsis thaliana*, and other plants and or mutants, variants, fragments or fusions of said SDFs and polypeptides or proteins derived therefrom. In some instances, the SDFs span the entirety of a protein-coding segment. In some instances, the entirety of an mRNA is represented. Other objects of the invention that are also represented by SDFs of the invention are control sequences, such as, but not limited to, promoters. Complements of any sequence of the invention are also considered part of the invention.

- 30 [0004] Other objects of the invention are polynucleotides comprising exon sequences, polynucleotides comprising intron sequences, polynucleotides comprising introns together with exons, intron/exon junction sequences, 5' untranslated sequences, and 3' untranslated sequences of the SDFs of the present invention. Polynucleotides representing the joinder of any exons described herein, in any arrangement, for example, to produce a sequence encoding any desirable amino acid sequence are within the scope of the invention.

- 35 [0005] The present invention also resides in probes useful for isolating and identifying nucleic acids that hybridize to an SDF of the invention. The probes can be of any length, but more typically are 12-2000 nucleotides in length; more typically, 15 to 200 nucleotides long; even more typically, 18 to 100 nucleotides long.

- 40 [0006] Yet another object of the invention is a method of isolating and/or identifying nucleic acids using the following steps:

- (a) contacting a probe of the instant invention with a polynucleotide sample under conditions that permit hybridization and formation of a polynucleotide duplex; and  
 (b) detecting and/or isolating the duplex of step (a).

- 45 [0007] The conditions for hybridization can be from low to moderate to high stringency conditions. The sample can include a polynucleotide having a sequence unique in a plant genome. Probes and methods of the invention are useful, for example, without limitation, for mapping of genetic traits and/or for positional cloning of a desired fragment of genomic DNA.

- 50 [0008] Probes and methods of the invention can also be used for detecting alternatively spliced messages within a species. Probes and methods of the invention can further be used to detect or isolate related genes in other plant species using genomic DNA (gDNA) and/or cDNA libraries. In some instances, especially when longer probes and low to moderate stringency hybridization conditions are used, the probe will hybridize to a plurality of cDNA and/or gDNA sequences of a plant. This approach is useful for isolating representatives of gene families which are identifiable by possession of a common functional domain in the gene product or which have common cis-acting regulatory sequences. This approach is also useful for identifying orthologous genes from other organisms.

- 55 [0009] The present invention also resides in constructs for modulating the expression of the genes comprised of all or a fragment of an SDF. The constructs comprise all or a fragment of the expressed SDF, or of a complementary

sequence. Examples of constructs include ribozymes comprising RNA encoded by an SDF or by a sequence complementary thereto, antisense constructs, constructs comprising coding regions or parts thereof, constructs comprising promoters, introns, untranslated regions, scaffold attachment regions, methylating regions, enhancing or reducing regions, DNA and chromatin conformation modifying sequences, etc. Such constructs can be constructed using viral, 5 plasmid, bacterial artificial chromosomes (BACs), plasmid artificial chromosomes (PACs), autonomous plant plasmids, plant artificial chromosomes or other types of vectors and exist in the plant as autonomous replicating sequences or as DNA integrated into the genome. When inserted into a host cell the construct is, preferably, functionally integrated with, or operatively linked to, a heterologous polynucleotide. For instance, a coding region from an SDF might be operably linked to a promoter that is functional in a plant.

10 [0010] The present invention also resides in host cells, including bacterial or yeast cells or plant cells, and plants that harbor constructs such as described above. Another aspect of the invention relates to methods for modulating expression of specific genes in plants by expression of the coding sequence of the constructs, by regulation of expression of one or more endogenous genes in a plant or by suppression of expression of the polynucleotides of the invention in a plant. Methods of modulation of gene expression include without limitation (1) inserting into a host cell additional 15 copies of a polynucleotide comprising a coding sequence; (2) modulating an endogenous promoter in a host cell; (3) inserting antisense or ribozyme constructs into a host cell and (4) inserting into a host cell a polynucleotide comprising a sequence encoding a variant, fragment, or fusion of the native polypeptides of the instant invention.

#### BRIEF DESCRIPTION OF THE TABLES

20 [0011] The sequences of exemplary SDFs and polypeptides corresponding to the coding sequences of the instant invention are described in Reference Tables 1 and 2, REF Tables 1 and 2"; and in Sequence Tables 1 and 2, SEQ Tables 1 and 2." The REF Tables refer to a number of Maximum Length Sequences" or MLS." Each MLS corresponds to the longest cDNA obtained, either by cloning or by the prediction from genomic sequence. The sequence of the 25 MLS is the cDNA sequence as described in the Av subsection of the REF Tables.

[0012] The REF Table includes the following information relating to each MLS:

I. cDNA Sequence

- 30 A. 5' UTR
- B. Coding Sequence
- C. 3' UTR

II. Genomic Sequence

- 35 A. Exons
- B. Introns
- C. Promoters

- 40 III. Link of cDNA Sequences to Clone IDs
- IV. Multiple Transcription Start Sites
- V. Polypeptide Sequences

- 45 A. Signal Peptide
- B. Domains
- C. Related Polypeptides

VI. Related Polynucleotide Sequences

50 I. cDNA SEQUENCE

[0013] The REF Tables indicate which sequence in the SEQ Tables represents the sequence of each MLS. The MLS sequence can comprise 5' and 3' UTR as well as coding sequences. In addition, specific cDNA clone numbers also are included in the REF Tables when the MLS sequence relates to a specific cDNA clone.

- 55 A. 5' UTR

[0014] The location of the 5' UTR can be determined by comparing the most 5' MLS sequence with the corresponding

## EP 1 033 405 A2

genomic sequence as indicated in the REF Tables. The sequence that matches, beginning at any of the transcriptional start sites and ending at the last nucleotide before any of the translational start sites corresponds to the 5' UTR.

### B. Coding Region

5

[0015] The coding region is the sequence in any open reading frame found in the MLS. Coding regions of interest are indicated in the Poly P SEQ subsection of the REF Tables.

### C. 3' UTR

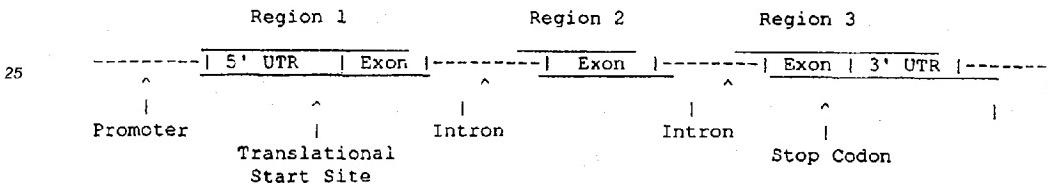
10

[0016] The location of the 3' UTR can be determined by comparing the most 3' MLS sequence with the corresponding genomic sequence as indicated in the REF Tables. The sequence that matches, beginning at the translational stop site and ending at the last nucleotide of the MLS corresponds to the 3' UTR.

## 15 II. GENOMIC SEQUENCE

[0017] Further, the REF Tables indicate the specific gi" number of the genomic sequence if the sequence resides in a public databank. For each genomic sequence, the REF Tables indicate which regions are included in the MLS. These regions can include the 5' and 3' UTRs as well as the coding sequence of the MLS. See, for example, the scheme below:

20



30

[0018] The REF Tables report the first and last base of each region that are included in an MLS sequence. An example is shown below:

35

gi No. 47000;

37102 ... 37497

37593 ... 37925

The numbers indicate that the MLS contains the following sequences from two regions of gi No. 47000; a first region including bases 37102-37497, and a second region including bases 37593-37925.

40

### A. EXON SEQUENCES

[0019] The location of the exons can be determined by comparing the sequence of the regions from the genomic sequences with the corresponding MLS sequence as indicated by the REF Tables.

45

#### i. INITIAL EXON

[0020] To determine the location of the initial exon, information from the

50

- (1) polypeptide sequence section;
- (2) cDNA polynucleotide section; and
- (3) the genomic sequence section

55

of the REF Tables are used. First, the polypeptide section will indicate where the translational start site is located in the MLS sequence. The MLS sequence can be matched to the genomic sequence that corresponds to the MLS. Based on the match between the MLS and corresponding genomic sequences, the location of the translational start site can be determined in one of the regions of the genomic sequence. The location of this translational start site is the start of the first exon.

[0021] Generally, the last base of the exon of the corresponding genomic region, in which the translational start site

[2347] The suspension culture cells are transformed with exogenous DNA as described by Z. Chen et al. *Plant Mol. Bio.* 36:163 (1998). Briefly, 4-days post-subculture cells are incubated with cell wall digestion solution containing 0.4 M sorbitol, 2% driselase, 5mM MES (2-[N-Morpholino] ethanesulfonic acid) pH 5.0 for 5 hours. The digested cells are pelleted gently at 60 xg for 5 min. and washed twice in W5 solution containing 154 mM NaCl, 5 mM KCl, 125 mM CaCl<sub>2</sub>, 0.5 M mannitol, pH 5.7 and the protoplast density is adjusted to about 4 x 10<sup>6</sup> protoplasts per ml.

[2348] 15-60 µg of plasmid DNA is mixed with 0.9 ml of protoplasts. The resulting suspension is mixed with 40% polyethylene glycol (MW 8000, PEG 8000), by gentle inversion a few times at room temperature for 5 to 25 min. Protoplast culture medium known in the art is added into the PEG-DNA-protoplast mixture. Protoplasts are incubated in the culture medium for 24 hour to 5 days and cell extracts can be used for assay of transient expression of the introduced gene. Alternatively, transformed cells can be used to produce transgenic callus, which in turn can be used to produce transgenic plants, by methods known in the art. See, for example, Nomura and Komamine, *Pit. Phys.* 79: 988-991 (1985), *Identification and Isolation of Single Cells that Produce Somatic Embryos in Carrot Suspension Cultures*.

[2349] The invention being thus described, it will be apparent to one of ordinary skill in the art that various modifications of the materials and methods for practicing the invention can be made. Such modifications are to be considered within the scope of the invention as defined by the following claims.

[2350] Each of the references from the patent and periodical literature cited herein is hereby expressly incorporated in its entirety by such citation.

20

### Claims

1. An isolated nucleic acid molecule comprising a nucleic acid having a nucleotide sequence which encodes an amino

25

acid sequence exhibiting at least 40% sequence identity to an amino acid sequence encoded by

- (a) a nucleotide sequence described in REF and/or SEQ Table 1 or 2 or a fragment thereof; or
- (b) a complement of a nucleotide sequence shown in REF and/or SEQ Table 1 or 2 or a fragment thereof.

30

2. An isolated nucleic acid molecule comprising a nucleic acid having a nucleotide sequence which exhibits at least 65% sequence identity to

- (a) a nucleotide sequence shown in REF and/or SEQ Table 1 or 2 or a fragment thereof; or
- (b) a complement of a nucleotide sequence shown in REF and/or SEQ Table 1 or 2 or a fragment thereof.

35

3. An isolated nucleic acid molecule comprising a nucleic acid having a nucleotide sequence which exhibits at least 65% sequence identity to a gene comprising

- (a) a nucleotide sequence shown in REF and/or SEQ Table 1 or 2 or a fragment thereof; or
- (b) a complement of a nucleotide sequence shown in REF and/or SEQ Table 1 or 2 or a fragment thereof.

40

4. An isolated nucleic acid molecule which is the reverse of the isolated nucleotide sequence according to any one of claims 1-3, such that the reverse nucleotide sequence has a sequence order which is the reverse of the sequence order of said isolated nucleotide sequence according to any one of claims 1-3.

45

5. An isolated nucleic acid molecule comprising a nucleic acid capable of hybridizing to a nucleic acid having a sequence selected from the group consisting of:

- (a) a nucleotide sequence which is shown in REF and/or SEQ Table 1 or 2; and
- (b) a nucleotide sequence which is complementary to a nucleotide sequence shown in REF and/or SEQ Table 1 or 2;

under conditions that permit formation of a nucleic acid duplex at a temperature from about 40°C and 48°C below the melting temperature of the nucleic acid duplex.

55

6. The nucleic acid molecule according to any one of claims 1-5, wherein said nucleic acid comprises an open reading frame.

7. The isolated nucleic acid molecule of any one of claims 1-5, wherein said nucleic acid is capable of functioning as a promoter, a 3' end termination sequence, an untranslated region (UTR), or as a regulatory sequence.
- 5 8. The isolated nucleic acid molecule of claim 7, wherein said nucleic acid is a promoter and comprises a sequence selected from the group consisting of a TATA box sequence, a CAAT box sequence, a motif of GCAATCG or any transcription-factor binding sequence, and any combination thereof.
- 10 9. The isolated nucleic acid molecule of claim 7, wherein the nucleic acid sequence is a regulatory sequence which is capable of promoting seed-specific expression, embryo-specific expression, ovule-specific expression, tapetum-specific expression or root-specific expression of a sequence or any combination thereof.
- 15 10. A vector construct comprising a nucleic acid molecule according to any one of claims 1-9, wherein said nucleic acid molecule is heterologous to any element in said vector construct.
11. A vector construct according to claim 10 comprising:
  - (a) a first nucleic acid having a regulatory sequence capable of causing transcription and/or translation; and
  - (b) a second nucleic acid having the sequence of said isolated nucleic acid molecule according to any one of claims 1-4;
- 20 wherein said first and second nucleic acids are operably linked and wherein said second nucleic acid is heterologous to any element in said vector construct.
12. The vector construct according to claim 11, wherein said first nucleic acid is native to said second nucleic acid.
- 25 13. The vector construct according to claim 11, wherein said first nucleic acid is heterologous to said second nucleic acid.
14. A vector construct according to claim 10 comprising:
  - (c) a first nucleic acid having the sequence of said isolated nucleic acid molecule according to claim 7; and
  - (d) a second nucleic acid;
- 30 wherein said first and second nucleic acids are operably linked and wherein said first nucleic acid is heterologous to any element in said vector construct.
15. The vector construct according to claim 14, wherein said first nucleic acid is native to said second nucleic acid.
- 40 16. The vector construct according to claim 14, wherein said first nucleic acid is heterologous to said second nucleic acid.
17. A host cell comprising an isolated nucleic acid molecule according to any one of claims 1-4, wherein said nucleic acid molecule is flanked by exogenous sequence.
- 45 18. A host cell comprising a vector construct of any one of claims 10-16.
19. An isolated polypeptide comprising an amino acid sequence
  - (a) exhibiting at least 40% sequence identity of an amino acid sequence encoded by a sequence shown in REF and/or SEQ Table 1 or 2 or a fragment thereof; and
  - (b) capable of exhibiting at least one of the biological activities of the polypeptide encoded by said nucleotide sequence shown in REF and/or SEQ Table 1 or 2 or a fragment thereof.
- 50 20. The isolated polypeptide of claim 19, wherein said amino acid sequence exhibits at least 75% sequence identity to an amino acid sequence encoded by a sequence shown in SEQ Table 1 or 2 or a fragment thereof.
21. The isolated polypeptide of claim 19, wherein said amino acid sequence exhibits at least 85% sequence identity

**EP 1 033 405 A2**

to an amino acid sequence encoded by a sequence shown in SEQ Table 1 or 2 or a fragment thereof.

22. The isolated polypeptide of claim 19, wherein said amino acid sequence exhibits at least 90% sequence identity to an amino acid sequence encoded by a sequence shown in SEQ Table 1 or 2 or a fragment thereof.

5

23. An antibody capable of binding the isolated polypeptide of any one of claims 19-22.

24. A method of introducing an isolated nucleic acid into a host cell comprising:

10

- (a) providing an isolated nucleic acid molecule according to any one of claims 1-4; and
- (b) contacting said isolated nucleic with said host cell under conditions that permit insertion of said nucleic acid into said host cell.

15

25. A method of transforming a host cell which comprises contacting a host cell with a vector construct according to any one of claims 10-16.

20

26. A method of modulating transcription and/or translation of a nucleic acid in a host cell comprising:

- (a) providing the host cell of claim 24 or 25; and
- (b) culturing said host cell under conditions that permit transcription or translation.

27. A method for detecting a nucleic acid in a sample which comprises:

25

- (a) providing an isolated nucleic acid molecule according to any one of claims 1-5;
- (b) contacting said isolated nucleic acid molecule with a sample under conditions which permit a comparison of the sequence of said isolated nucleic acid molecule with the sequence of DNA in said sample; and
- (c) analyzing the result of said comparison.

30

28. The method according to claim 27, wherein said isolated nucleic acid molecule and said sample are contacted under conditions which permit the formation of a duplex between complementary nucleic acid sequences.

35 29. A plant or cell of a plant which comprises a nucleic acid molecule according to any one of claims 1-4 which is exogenous to said plant or plant cell.

30

30. A plant or cell of a plant which comprises a nucleic acid molecule according to any one of claims 1-4, wherein said nucleic acid molecule is heterologous to said plant or said cell of a plant.

35 31. A plant or cell of a plant which has been transformed with a nucleic acid molecule according to any one of claims 1-4.

40

32. A plant or cell of a plant which comprises a vector construct according to any one of claims 10-16.

33. A plant or cell of a plant which has been transformed with a vector construct according to any one of claims 10-16.

45

34. A plant which has been regenerated from a plant cell according to any one of claims 29-33.

50

55